**Section – A (20 questions)**

**1. What is the fundamental difference between Artificial Intelligence (AI) and Machine Learning (ML)?**

a) AI requires human intervention, while ML does not.

b) AI focuses on mimicking human intelligence, while ML focuses on learning from data.

c) AI is limited to rule-based systems, while ML can handle complex tasks.

d) AI and ML are synonymous terms.

**2. Which of the following is NOT a basic data structure in Python?**

a) List b) Set c) Queue d) Tuple

**3. What method is used to read a CSV file into a Pandas DataFrame?**

a) `read\_csv()` b) `load\_csv()` c) `import\_csv()` d) `open\_csv()`

**4. Which statistical measure provides information about the variability or spread of data?**

a) Mean b) Median c) Mode d) Standard deviation

**5. Which EDA technique is used to identify relationships between two numerical variables?**

a) Histogram b) Box plot c) Scatter plot d) Bar plot

**6. Which Python library is best suited for creating static, publication-quality plots?**

a) Matplotlib b) Seaborn c) Plotly d) ggplot

**7. Which type of learning involves providing the model with labeled data for training?**

a) Supervised learning b) Unsupervised learning

c) Semi-supervised learning d) Reinforcement learning

**8. Which of the following is NOT a machine learning algorithm available in scikit-learn?**

a) Decision Tree b) Support Vector Machine (SVM)

c) k-means Clustering d) Linear Regression

**9. What is the primary characteristic of supervised learning?**

a) Requires labeled data for training b) Learns from unlabeled data

c) Does not require any input data d) Focuses on mimicking human intelligence

**10. Which library provides high-level mathematical functions to operate on arrays and matrices?**

a) Matplotlib b) Pandas c) NumPy d) SciPy

**11. What method is used to do slicing in Pandas DataFrame?**

a) `get\_columns()` b) `select\_columns()` c) `filter()` d) `iloc[]`

**12. Which measure of central tendency is not affected by extreme values or outliers?**

a) Mean b) Median c) Mode d) Variance

**13. Which EDA technique is used to identify missing values in a dataset?**

a) Histogram b) Box plot c) Scatter plot d) Heatmap

**14. Which Python library is specifically designed for statistical data visualization based on Matplotlib?**

a) Seaborn b) Plotly c) ggplot d) Bokeh

**15. In unsupervised learning, what is the primary objective?**

a) Predict a target variable b) Discover patterns or structures in data

c) Classify data into predefined categories d) Minimize errors between predicted and actual values

**16. Which of the following is a classification algorithm available in scikit-learn?**

a) K-means Clustering b) Decision Tree

c) Principal Component Analysis (PCA) d) Linear Regression

**17. What is the main difference between supervised and unsupervised learning?**

a) Supervised learning requires labeled data, while unsupervised learning does not.

b) Unsupervised learning requires labeled data, while supervised learning does not.

c) Supervised learning is faster than unsupervised learning.

d) Unsupervised learning focuses on classification tasks, while supervised learning focuses on clustering tasks.

**18. Which Python library provides tools for data manipulation and analysis, especially when working with structured data?**

a) NumPy b) Pandas c) Matplotlib d) Scikit-learn

**19. Which method is used to drop duplicate rows from a Pandas DataFrame?**

a) `drop\_duplicates()` b) `remove\_duplicates()` c) `drop\_rows()` d) `remove\_rows()`

**20. Which measure of dispersion represents the average deviation of each data point from the mean?**

a) Variance b) Standard deviation c) Range d) Mean absolute deviation (MAD)

Section – B (15 question)

**21. What is the key challenge in unsupervised learning compared to supervised learning?**

a) Lack of computational resources b) Difficulty in defining the target variable

c) Limited availability of labeled data d) Inability to handle non-linear relationships

**22. In Python, what is the purpose of the `map` function?**

a) Apply a function to each element of an iterable b) Create a mapping of key-value pairs

c) Filter elements from an iterable d) Generate a sequence of random numbers

**23. Consider a DataFrame 'df' with columns 'A' and 'B'. How can you drop all rows where column 'A' is greater than 10?**

a) `df.drop(df['A'] > 10)` b) `df.drop(df[df['A'] > 10].index)` c) `df.drop\_rows('A > 10')`

d) `df.filter(df['A'] > 10)`

**24. What is the purpose of the Shapiro-Wilk test in statistics?**

a) Test for normality of a distribution b) Compare means of two independent samples

c) Assess correlation between variables d) Determine homogeneity of variances

**25. In dimensionality reduction, what is the primary goal of Principal Component Analysis (PCA)?**

a) Increase interpretability of data b) Identify outliers in the dataset

c) Reduce the number of features while retaining information

d) Enhance visualization of high-dimensional data

26. **Which type of plot is suitable for visualizing the relationship between two continuous variables and their joint distribution?**

a) Line plot b) Scatter plot c) Histogram d) Box plot

**27. What is the Curse of Dimensionality in the context of machine learning?**

a) The difficulty of defining target variables in high-dimensional data

b) The computational burden of handling large datasets

c) The impact of irrelevant features on model performance**.**

d) The deterioration of algorithm performance as the number of features increases**.**

**28. What is the purpose of cross-validation in machine learning?**

a) Test the model on a completely new dataset

b) Optimize hyperparameters for better performance

c) Evaluate model performance using a single split of data

d) Mitigate the risk of overfitting by using multiple train-test splits.

**29. Which type of learning algorithm is most susceptible to overfitting?**

a) Decision Tree b) k-Nearest Neighbors (KNN)

c) Support Vector Machine (SVM) d) Naive Bayes

**30. What does the term "Pythonic" mean in the context of Python programming?**

a) Code that follows strict PEP 8 guidelines b) Code that is written in Python 2.x syntax

c) Code that adheres to industry-standard conventions

d) Code that embraces the idioms and conventions of the Python language

**31. Given a DataFrame 'df' with columns 'Category' and 'Value', how can you calculate the mean 'Value' for each 'Category'?**

a) `df.group\_by('Category').mean()` b) `df.groupby('Category')['Value'].mean()`

c) `df.mean('Category', 'Value')` d) `df.calculate\_mean('Category', 'Value')`

**32. What is the purpose of the skewness measure in descriptive statistics?**

a) Assess the symmetry of a distribution b) Measure the spread of data

c) Determine the presence of outliers d) Quantify the central tendency of a dataset

**33. In feature scaling, what is the objective of transforming numerical variables to have a mean of 0 and a standard deviation of 1?**

a) Enhance interpretability of the data b) Facilitate visualization of data

c) Normalize variables with different scales d) Reduce the impact of outliers on the analysis

**34. What is the primary purpose of a violin plot in data visualization?**

a) Display the distribution of a single numerical variable

b) Show the relationship between two numerical variables

c) Compare the means of different categories d) Provide a summary of statistical measures.

**35.Which algorithm is commonly used for regression tasks in machine learning?**

a) k-Means Clustering b) Random Forest

c) k-Nearest Neighbors (KNN) d) Support Vector Machine (SVM)